AD-A044 574 GEORGE WASHINGTON UNIV WASHINGTON D C INST FOR MANAG--ETC F/G 12/1
AN IDEALIZED EXACT PENALTY FUNCTION.(U)
JUL 77 G P MCCORMICK AFOSR-73-2504
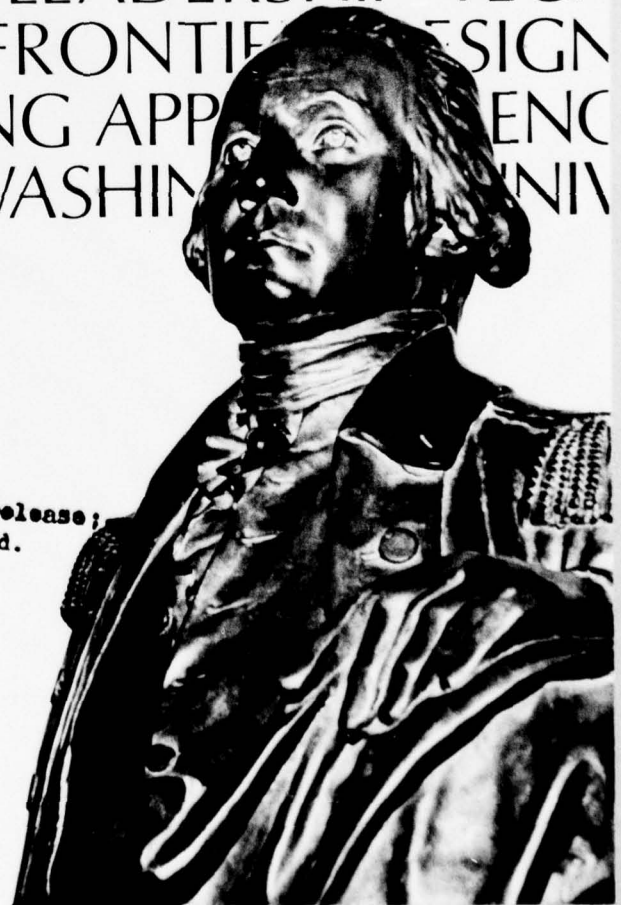UNCLASSIFIED SERIAL-T-359 AFOSR-TR-77-1213 NL

AD
A044574

END
DATE
FILMED
10-77
DDC

AFOSR-TR- 77- 1213

ADA044574

STUDENTS FACULTY STUDY R
ESEARCH DEVELOPMENT FUT
URE CAREER CREATIVITY CC
MMUNITY LEADERSHIP TECH
NOLOGY FRONTIE ESIGN
ENGINEERING APP ENC
GEORGE WASHIN NIV

Approved for public release;
distribution unlimited.

INSTITUTE FOR MANAGEMENT
SCIENCE AND ENGINEERING

SCHOOL OF ENGINEERING
AND APPLIED SCIENCE

AN IDEALIZED EXACT PENALTY FUNCTION

by

Garth P. McCormick

Serial T-359
14 July 1977

The George Washington University
School of Engineering and Applied Science
Institute for Management Science and Engineering

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Abstract
of
Serial T-359
14 July 1977

AN IDEALIZED EXACT PENALTY FUNCTION

by

Garth P. McCormick

In this paper an idealized exact penalty function is derived from natural considerations of the flow of particles under different forces. It is shown how Fletcher's exact penalty function is an approximation to this one. A second order version of the idealized exact penalty function is developed which is computable.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-77-1313	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN IDEALIZED EXACT PENALTY FUNCTION		5. TYPE OF REPORT & PERIOD COVERED Interim rept.
		6. PERFORMING ORG. REPORT NUMBER Serial-T-359
7. AUTHOR(s) Garth P. McCormick		8. CONTRACT OR GRANT NUMBER(s) AFOSR-73-2504
9. PERFORMING ORGANIZATION NAME AND ADDRESS The George Washington University School of Engineering & Applied Science Washington, DC 20052		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A6
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE 14 Jul 77
		13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 32p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

exact penalty function
nonlinear programming

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper an idealized exact penalty function is derived from natural considerations of the flow of particles under different forces. It is shown how Fletcher's exact penalty function is an approximation to this one. A second order version of the idealized exact penalty function is developed which is computable.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE   UNCLASSIFIED

406 743

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

AN IDEALIZED EXACT PENALTY FUNCTION

by

Garth P. McCormick

1. Introduction

In [1] and [2] Fletcher developed an exact penalty function for constrained optimization problems. That is, he showed how one unconstrained minimization problem could yield the solution of a constrained problem (in a sense to be defined in Section 2). In this paper, an idealized exact penalty function is derived from considerations of the movement of a particle under different forces. This is done first for the equality constrained problem and then generalized to the inequality constrained case.

The idealized exact function has flow lines similar to those observed for the particle. It is shown that Fletcher's exact penalty function is an approximation to the idealized one near constraint boundaries. A new computable exact penalty function which uses second order information is developed which provides a better approximation to the idealized one.

2. Movement of a Particle Under Different Forces

Consider the equality constrained nonlinear programming problem:

$$\text{minimize} \quad f(x)$$
$$x \in E^n$$

$$\text{subject to} \quad x \in R = \{x \mid h(x) = 0\} \tag{1}$$

where h(x) is a p by one vector of functions and where $x \in E^n$ is
an n dimensional vector.

A physical situation which models this is one in which the particle
in E^n is acted upon by forces which tend to decrease the function f(x)
and which also tend to drive the particle to a point satisfying the con-
straints. Many models of flow result depending upon the relative magnitudes
of the force lines. In Figure 1 is one situation associated with the problem

$$\text{minimize} \quad -x + y$$
$$(x,y) \tag{2}$$
$$\text{subject to} \quad x^2 + y^2 = 1 .$$

In Figure 1 the lines of force off the perimeter of the circle are entirely
associated with driving the particle to a feasible point. This is done in
a way to minimize the distance between the particle and the feasible region.
Once on perimeter, the lines of force become the gradient of the objective
function projected onto the tangent plane.

An exact penalty function whose gradient vector approximates this
flow pattern can be formulated as:

$$E(x) = \begin{cases} f[Pr(x)] & \text{if } x \in R \\ \mathscr{D}(x) & \text{if } x \notin R \end{cases} \tag{3}$$

where Pr(x) is the projection function which maps any point x into the
'closest feasible point' and where $\mathscr{D}(x)$ is the minimum distance from x
to the set R given by

$$\mathscr{D}(x) = \min_{y \in R} || x - y || . \tag{4}$$

The major difficulty with this exact penalty function is that neither
it nor its derivative is continuous. Algorithms for minimizing unconstrained
functions require (usually) that the first derivatives of the function be
continuous. Furthermore, following these flow lines is not necessarily the
quickest way to solve the problem. A better strategy would be to anticipate

MINIMIZE $-x + y = f$
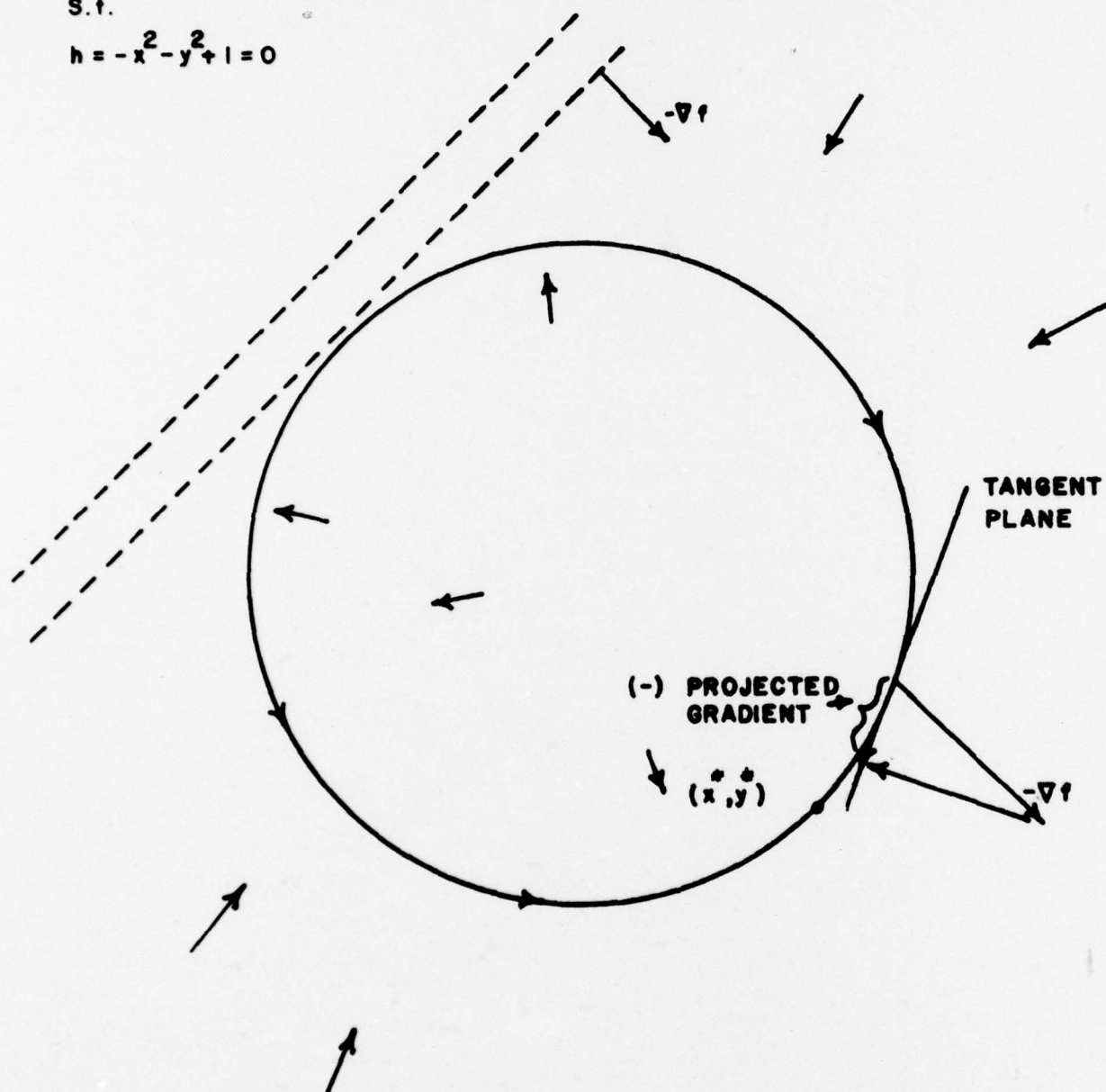S.t.
$h = -x^2 - y^2 + 1 = 0$

FIGURE I   PARTICLE FLOW LINES

the flow near the perimeter and create a line of flow which is a combination of the projected gradient on the perimeter and the gradient of the squared distance function. The resulting lines of flow would be those pictured in Figure 2. An associated natural exact penalty function would be

$$E(x) = f[x-d(x)] + qd(x)^T d(x) \tag{5}$$

where $x - d(x)$ is a point solving (4) and q is a scalar greater than 0. There are difficulties with this definition involving uniqueness. Note that $d(x)$ is not unique in example (2) when $(x,y) = (0,0)$. Usually the vector $d(x)$ is unique but if it is not, to complete the definition, the following is used:

$$\text{define } D(x) = \{d \mid x-d \text{ solves (4)}\}.$$

Let $d(x)$ be a vector from $D(x)$ such that

$$f[x-d(x)] \tag{6}$$

is minimal. This specifies the value of the function $E(x)$ although not necessarily the vector $d(x)$ since theoretically there can be more than one vector in $D(x)$ satisfying the criterion (6).

This exact penalty function (5) combines a penalty associated with being away from the feasible region and the value of the objective function at the closest feasible point. The differentiability of the function (5) depends upon the differentiability of the distance function $d(x)$.

Some isovalue contours of this penalty function associated with problem (2) when $q = 1$ are given in Figure 3. Note that this is not continuous at $(0,0)$, although from the definition $E(0,0) = -\sqrt{2} + (1/4)$ since $D(0,0) = \{d \mid ||d||^2 = 1\}$ and $d(0,0) = (+\sqrt{2}/2, -\sqrt{2}/2)$.

The equivalence of the unconstrained minimization of the idealized exact penalty function (5) and the constrained problem (1) is summarized in the next three theorems.
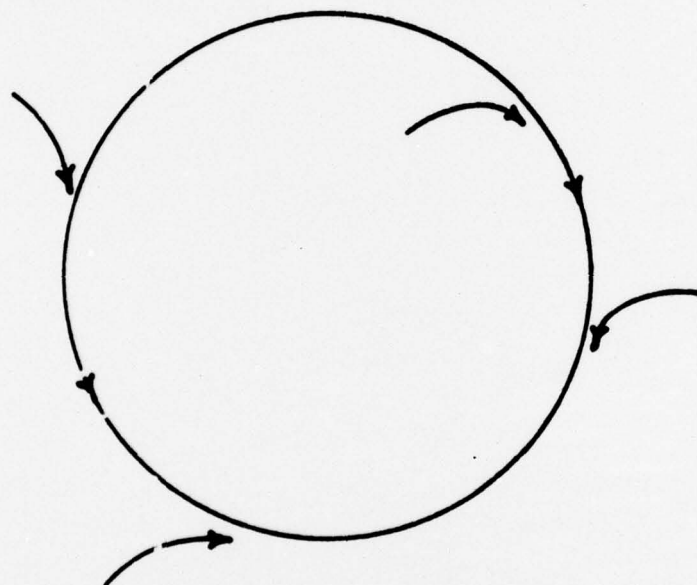
FIGURE 2   FLOW OF PARTICLE WITH
            ANTICIPATION OF BOUNDARY

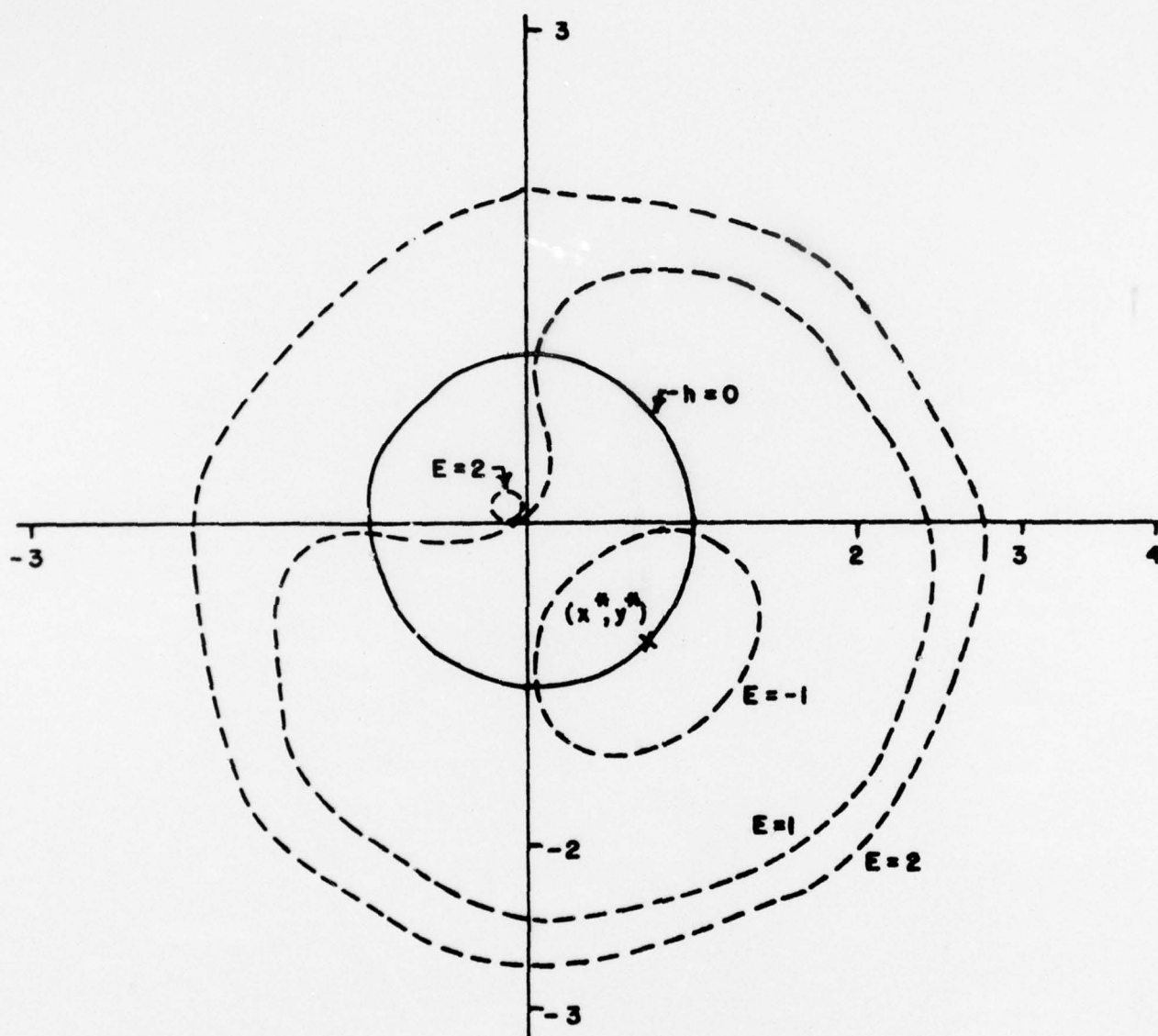$$E(x,y) = -x/(x^2+y^2)^{.5} + y/(x^2+y^2)^{.5} + [(x^2+y^2)^{.5}-1]^2$$

FIGURE 3 ISOVALUE CONTOURS FOR
EXACT PENALTY FUNCTION

Theorem 1. (Assume f, $\{h_j\}$ are continuous.) If $\bar{x}$ is a local uncon-strained minimizer for (5), it is a local minimizer for (1).

Proof: First it will be shown that if $d(x) \neq 0$, for every $0 < \varepsilon \leq 1$,

$$E[x-\varepsilon d(x)] < E(x) . \tag{7}$$

Obviously $d(x) \neq 0 \Rightarrow d[x-\varepsilon d(x)] \neq 0$, since otherwise $x - \varepsilon d(x)$ would be a closer feasible point to x than $x - d(x)$. Also note that

$$[x-\varepsilon d(x)] - (1-\varepsilon)d(x) \; \varepsilon \; \bar{R} \; ,$$

and thus by definition

$$||d[x-\varepsilon d(x)]|| \leq (1-\varepsilon)||d(x)|| . \tag{8}$$

Also by definition,

$$||d(x)|| \leq ||x - \{[x-\varepsilon d(x)]-d[x-\varepsilon d(x)]\}||$$
$$\tag{9}$$
$$< \varepsilon||d(x)|| + ||d[x-\varepsilon d(x)]|| \qquad \text{(using the triangular inequality)}$$

(assuming $d(x)$ is not proportional to $d[x-\varepsilon d(x)]$) . Together (8) and (9) imply that $(1-\varepsilon)d(x)$ is the single element in $D[x-\varepsilon d(x)]$. Because of (6),

$$f[x-\varepsilon d(x)-d[x-\varepsilon d(x)]] + q||d[x-\varepsilon d(x)]||^2 = f[x-d(x)] + q(1-\varepsilon)^2||d(x)||^2$$
$$< f[x-d(x)] + q||d(x)||^2 \quad \text{(because } d(x) \neq 0 \text{ assumed).}$$

This completes the proof of (7).

Because $\bar{x}$ is a local unconstrained minimizer it follows from (7) that $d(\bar{x}) = 0$ and hence that $f(\bar{x}) = E(\bar{x})$. Furthermore, there is a neigh-borhood $N(\bar{x},\delta)$ such that $E(x) \geq E(\bar{x})$ for all $x \; \varepsilon \; N(\bar{x},\delta)$. Let $x \; \varepsilon \; R \cap N(\bar{x},\delta)$.

- 7 -

Then

$$f(x) = E(x) \qquad (d(x) = 0 \quad \text{for} \quad x \ \varepsilon \ R)$$

$$\geq E(\overline{x}) \qquad (\text{since} \quad x \ \varepsilon \ N(\overline{x},\delta) \ .$$

$$= f(\overline{x})$$

This completes the proof of the theorem.

Theorem 1 is important in that it states that use of the idealized exact penalty function does not introduce spurious minimizers. In the next theorem the converse is proved.

Theorem 2. Suppose f, $\{h_j\}$ are continuous functions. If $\overline{x}$ is a local minimizer for (1), then it is an unconstrained local minimizer for (5).

Proof: If the theorem is not true, there is an infinite sequence of points $\{x_k\}$ where $x_k \to \overline{x}$ such that

$$f[x_k - d(x_k)]$$

$$\leq f[x_k - d(x_k)] + qd(x_k)^T d(x_k) \equiv E(x_k)$$

$$< E(\overline{x}) = f(\overline{x}) \ ,$$

for all k, with $x_k - d(x_k) \ \varepsilon \ R$. Since $x_k - \overline{x} \to 0$, $d(x_k) \to 0$, and therefore $x_k - d(x_k) \to \overline{x}$. This contradicts the assumption that $\overline{x}$ is a minimizer for $x \ \varepsilon \ R$.

Theorem 3. Suppose f, $\{h_j\}$ are continuous functions. If $\overline{x}$ is a global unconstrained minimizer for (5), it is a global minimizer for (1). If x^* is a global minimizer for (1), it is a global unconstrained minimizer for (5).

Proof: The proof is obvious and will not be given.

It is useful to examine the idealized exact penalty function in terms of its derivative and its Hessian matrix. These results will be useful in developing a computable exact penalty function in the next section.

__Theorem 4.__ Suppose f, $\{h_j\} \in C^3$. Suppose $\bar{x}$ is a point where $f'(\bar{x})$ has full row rank and suppose further that $d(x)$ is unique and continuously differentiable in a neighborhood about $\bar{x}$. If $\bar{x}$ is an isolated unconstrained minimizer for the idealized penalty function (5), i.e., if $E'(\bar{x}) = 0$ and $E''(\bar{x})$ is a positive definite matrix, then $\bar{x}$ satisfies the second order sufficiency conditions for an isolated local minimizer for (1).

__Proof:__ Assume for the moment that at a point x , $d'(x)$ and $d''(x)$ are defined. Using the chain rule of differentiation,

$$E'(x) = f'[x-d(x)][I-d(x)] + 2q\, d(x)^T d'(x) , \qquad (10)$$

and

$$E''(x) = \{f'[x-d(x)] \otimes I_n\} \{-d''(x)\} + [I-d'(x)]\, f''[x-d(x)]\, [I-d'(x)]$$
$$+ 2q\, d'(x)^T d'(x) + 2q[d(x) \otimes I_n]d''(x) . \qquad (11)$$

Consider any point x near $\bar{x}$. Let $z^*(x)$ solve the problem

$$\underset{z}{\text{minimize}} \ ||z-x||^2 \ \text{subject to} \ h(z) = 0 .$$

Note that $d(x) = x - z^*(x)$. For x close enough to $\bar{x}$, clearly $z^*(x)$ is close to $\bar{x}$ and the matrix $f'[z^*(x)]$ has rank p . Therefore, the first order necessary conditions apply and

$$[z^*(x)-x]\, 2 + h'[z^*(x)]^T\, u^*(x) = 0 ,$$

where

$$u^*(x) = - \{h'[z^*(x)]h'[z^*(x)]^T\}^{-1} h'[z^*(x)]\, [x-z^*(x)]2 .$$

Another way of writing this is

$$P(x)d(x) = 0 \qquad (12)$$

where

$$P(x) \equiv \left[I - h'[z^*(x)]^T \{h'[z^*(x)]h'[z^*(x)]^T\}^{-1} h'[z^*(x)] \right] .$$

- 9 -

Also,

$$h[x-d(x)] = h[z*(x)] = 0 . \tag{13}$$

Differentiating (13) yields

$$h'[x-d(x)] \, [I-d'(x)] = 0 . \tag{14}$$

Differentiating (12) yields

$$P(x)d'(x) + [I_n \otimes d(x)^T]P'(x) = 0 . \tag{15}$$

When $d(x) = 0$, (14) implies that

$$h'(x) = h'(x)d'(x)$$

and using this in (15) yields

$$d'(x) = h'(x)\{h'(x)h'(x)^T\}h'(x)^T . \tag{16}$$

Differentiation of (14) directly yields

$$\{h_j'[z*(x)] \otimes \} \{-d''(x)\} + \{I-d'(x)\}h_j''[z*(x)]\{I-d'(x)\} = 0 . \tag{17}$$

From Theorem 2 it is known that $d(\bar{x}) = 0$ (and therefore that $h(\bar{x}) = 0$). Thus, formula (16) can be used.

Then $E'(\bar{x}) = 0$, implies, using (10) and (16), that the first order necessary conditions are satisfied at $\bar{x}$ for a constrained minimizer. The appropriate Lagrange multipliers are given by the formula

$$u(\bar{x})^T = f'(\bar{x})h'(\bar{x})^T\{h'(\bar{x})h'(\bar{x})^T\}^{-1} .$$

Thus,

$$f'(\bar{x}) = u(\bar{x})h'(\bar{x}) . \tag{18}$$

The first term in (11) can be replaced using (17) and (18) and summing appropriately with

$$- [I-d'(\bar{x})] \, [\{u(\bar{x})^T \otimes I_n\}h''(\bar{x})] \, [I-d'(\bar{x})] .$$

- 10 -

Using this, then,

$$E''(\bar{x}) = P(\bar{x})\left[f''(\bar{x})-\{u(\bar{x})^T \otimes I_n\}h''(\bar{x})\right]P(\bar{x}) + 2qh'(\bar{x})^T[h'(\bar{x})h'(\bar{x})^T]^{-1}h'(\bar{x}) \quad . \quad (19)$$

Because $E''(\bar{x})$ was assumed positive definite, it follows that

$$z^T P(\bar{x})\left[f''(\bar{x})-\{u(\bar{x})^T \otimes I_n\}h''(\bar{x})\right]P(\bar{x})z = z^T[f''(\bar{x}) - \{u(\bar{x})^T \otimes I_n\}h''(\bar{x})]z > 0$$

for all z where $h'(\bar{x})z = 0$. Thus the second order sufficiency conditions
are satisfied at $\bar{x}$.

3. Fletcher's Exact Penalty Function (Equality Case)

In [1], Fletcher proposed an exact penalty function (with variations)
for the equality constrained optimization problem (1). The variation closest
to the natural idealized function developed in (5) is

$$F(x) = f(x) - f'(x) h'(x)^{+} h(x) + qh(x)^T h'(x)^{+T} h'(x)^{+} h(x) \quad\quad\quad (20)$$

where $h'(x)^{+}$ is the Penrose-Moore generalized inverse of $h'(x)$, the p by
n derivative matrix of $h(x)$. In general, for a matrix A , A^{+} is the
unique matrix satisfying

$$AA^{+}A = A$$
$$A^{+}AA^{+} = A^{+}$$
$$(AA^{+})^T = AA^{+}$$
$$(A^{+}A)^T = A^{+}A \quad .$$

Let $z^*(x)$ denote a solution of the minimum distance problem:

$$\underset{z}{\text{minimize}} \; || x-z ||^2$$

$$\text{subject to} \; h(z) = 0 \quad .$$

A first order Taylor's series approximation yields

$$0 = h[z^*(x)] \doteq h(x) + h'(x)[z^*(x)-x] \ .$$

The solution to this approximation with miniumum norm is

$$- d(x) = z^*(x)-x = - h'(x)^\dagger h(x) \ .$$

Substituting this in the idealized exact penalty function (5) using the approximation $f[x-d(x)] \doteq f(x) - f'(x)d(x)$ yields (20) above directly.

Viewed from this point of view, Fletcher's exact penalty function for equality constraints is a first-order approximation to the idealized penalty function (5).

Fletcher was able to show that if a point x^* satisfied the second order sufficiency conditions associated with (1), it was an isolated local unconstrained minimizer of the penalty function (10) for a large enough value of q . This is a weaker theorem than Theorem 2 which made that statement for any value of q . More important, there is no corresponding theorem for (20) analogous to Theorem 1. That is, the question of whether or not (20) has local unconstrained minimizers which are not local minimizers for the equality constrained problem (1) was not resolved. A partial resolution of this question can be obtained by making a more precise approximation to $z^*(x)$.

Using a second order approximation,

$$- d(x) = z^*(x) - x \doteq - h'(x)^\dagger [h(x) + \frac{1}{2} \gamma(x)]$$

where $\gamma(x)$ is a p by 1 vector whose jth component is

$$\gamma_j(x) = h(x)^T h'(x)^{\dagger T} h''_j(x) h'(x)^\dagger h(x) \ .$$

Using the second order approximation

$$f[x-d(x)] \doteq f(x) - f'(x)d(x) + \frac{1}{2} d(x)^T f''(x)d(x)$$

and substituting the $d(x)$ above in (5) yields (throwing away terms beyond quadratic in $h(x)$)

$$M(x) = f(x) - f'(x) \, h'(x)^{\dagger} \, h(x) - \frac{1}{2} f'(x) h'(x)^{\dagger} \, \gamma(x)$$

$$+ \frac{1}{2} h(x)^{T} \, h'(x)^{\dagger T} \, f''(x) \, h'(x)^{\dagger} \, h(x) \tag{21}$$

$$+ q h(x) h'(x)^{\dagger T} \, h'(x)^{\dagger} \, h'(x) \ .$$

When at x the derivatives $\{h_j'(x)^{T}\}$, $j = 1,\ldots,p$ are linearly independent, then

$$h'(x)^{\dagger} = h'(x)^{T} \, \{h'(x) h'(x)^{T}\}^{-1} \ .$$

The quantity

$$u(x) = h'(x)^{\dagger T} \, f'(x)^{T}$$

is an estimate of the Lagrange multipliers usually associated with a local minimizer of (1). In order to differentiate $M(x)$ is it necessary to obtain $u'(x)$. When $h'(x)$ has full row rank, it can be shown that

$$u'(x) = [h'(x) h'(x)^{T}]^{-1} \left[\{I_p \otimes [f'(x) - u(x)^{T} h'(x)]\} \, h''(x) \right]$$

$$\tag{22}$$

$$+ h'(x)^{\dagger T} \left[f''(x) - \{u(x)^{T} \otimes I_n\} \, h''(x) \right] \ .$$

Assume that at a point x, $h'(x)$ has full row rank and that $h(x) = 0$. Using (22), then $M'(x)$ and $M''(x)$ are

$$M'(x) = f'(x) - u(x)^{T} h'(x) \ , \tag{23}$$

and

$$M''(x) = - h'(x)^{T} \left[\{I_p \otimes [f'(x) - u(x)^{T} h'(x)]\} \, h''(x) \right]$$

$$- \left[I_p \otimes \{f'(x) - u(x)^{T} h'(x)\} h''(x) \right]^{T} h''(x)^{\dagger T}$$

$$\tag{24}$$

$$+ h'(x)^{T} \left[h'(x) \, h'(x)^{T} \right]^{-1} h'(x) 2q$$

$$+ P(x) \left[f''(x) - \{u(x)^{T} \otimes I_n\} \, h''(x) \right] P(x)$$

- 13 -

where

$$P(x) \equiv \left[I - h'(x)^{+}h'(x)\right] = \left[I - h'(x)^{T}\{h'(x)h'(x)^{T}\}^{-1}h'(x)\right] .$$

<u>Theorem 5</u>. Suppose f, $\{h_j\} \in C^2$. Let $\bar{x}$ be an unconstrained local minimizer for the exact penalty function $M(x)$ given by (21). If $h(\bar{x}) = 0$, then $\bar{x}$ is a constrained minimizer for problem (1).

<u>Proof</u>: Let $x \in R$ be any point "close to" $\bar{x}$.

Then

$$f(x) = M(x) \quad (x \in R \text{ and therefore } h(x) = 0)$$

$$\geq M(\bar{x}) \quad (\bar{x} \text{ is a local unconstrained minimizer})$$

$$= f(\bar{x}) \quad (\bar{x} \in R \text{ and therefore } h(\bar{x}) = 0) .$$

The same theorem under the same assumptions can be proved for Fletcher's penalty function (20).

<u>Theorem 6</u>. Suppose f , $\{h_j\} \in C^2$. Let $\bar{x}$ be an unconstrained local minimizer for the exact penalty function $F(x)$ given by (20). If $h(\bar{x}) = 0$, then $\bar{x}$ is a local minimizer for the constrained problem (1).

<u>Proof</u>: The proof is identical to that of Theorem 5.

The difference between these approximations and the idealized exact penalty function is that they do not guarantee, except in special cases, that the penalty function is the value of the objective function at the closest feasible point, plus some weight of the squared distance to that point. When the constraints are linear and the objective function quadratic, $M(x)$ does have this property and, therefore, one can show that unconstrained local minimizers in this instance are constrained local minimizers (actually global minimizers). Since $F(x)$ is a first-order approximation to (5) and $M(x)$ is a second-order approximation, one would expect that examples where spurious local unconstrained minimizers exist to the former which do not correspond to constrained minimizers of (1) would be more unlikely to occur for $M(x)$. At present there are no examples of this.

From an algorithmic point of view, if in attempting to minimize either $F(x)$ or $M(x)$, the sequence of points generated is not tending toward feasiblity, it would be no great trick to modify the algorithm to obtain points closer to feasibility and then retry to minimize the exact penalty functions later.

The next theorem shows another sense in which $M(x)$ is closer to the idealized exact penalty function. The Hessian matrix of $M(x)$ at unconstrained local minimizers for (1) agrees with that of $E(x)$. This is not the case for $F(x)$ which is why the value of q is impoitant for showing that strict local minimizers of (1) are isolated unconstrained minimizers for $F(x)$.

Theorem 7. Suppose f, $\{h_j\} \in C^3$. Suppose $\bar{x}$ is a point where $f'(\bar{x})$ has full row rank, and suppose further that $\bar{x}$ satisfies the second-order sufficiency conditions for a strict isolated local minimizer. Then $\bar{x}$ is an isolated unconstrained local unconstrained minimizer for $M(x)$ as given by (21) for any value of $q > 0$.

Proof: Because $\bar{x}$ is feasible, $h(\bar{x}) = 0$, and (23) and (24) are applicable. From the first-order optimality conditions for a constrained minimizer, it is known that

$$f'(\bar{x}) - u(\bar{x})^T h'(\bar{x}) = 0 , \tag{25}$$

therefore, $M'(\bar{x}) = 0$.

Because of (25), $E''(\bar{x})$ takes the form

$$P(\bar{x})\left[f''(\bar{x}) - \{u(\bar{x})^T \otimes I_n\}h''(\bar{x})\right]P(\bar{x}) + 2qh'(\bar{x})^T\{h'(\bar{x})h'(\bar{x})^T\}^{-1}h'(x) .$$

The second order sufficiency conditions imply that this is positive definite for every $q > 0$ and therefore that $\bar{x}$ is an isolated unconstrained local minimizer for (21).

4. The Inequality Constrained Problem

The inequality constrained problem can be written:

$$\text{minimize} \quad f(x)$$
$$x$$

$$\text{subject to} \quad g_i(x) \geq 0 \text{ , for } i = 1,\ldots,m \text{ .}$$

(26)

Define $R = \{x \,|\, g_i(x) \geq 0 \text{ , } i = 1,\ldots,m\}$.

An example of this problem is a modification of Problem (2):

$$\text{minimize} \quad -x + y$$
$$(x,y)$$

$$\text{subject to} \quad -x^2 - y^2 + 1 \geq 0 \text{ .}$$

(27)

Lines of flow associated with this problem can take forms similar to the equality constrained Problem (2) with the major exception that the interior of the circle is also now feasible. This means (see Figure 1) the flow lines in the circle which are above the line $-x + y = 0$ are no longer valid. They would be replaced by lines parallel to the negative gradient vector $(1,-1)^T$. In Figure two, the modified lines of flow would probably look like those given in Figure 4. Here the lines would follow the negative gradient path (in the interior of the circle) until the boundary is "sensed" and then would curve as a combination of the projected gradient at the boundary and the negative gradient vector.

A useful way to view this modification is to decompose the negative gradient of f into two parts. One is the projection of the negative gradient on the direction which tends to the closest boundary point, and the other that which is the difference of the negative gradient and the projected negative gradient. This decomposition is shown in Figure 5. A natural modification of the lines of flow is then to truncate the projected negative gradient vector if it extends beyond the feasible region.

This analysis of the lines of force makes clear the exact penalty function to be used in circumstances near constraint boundaries where the objective function has a lesser value at the closest point. The analysis is dependent upon
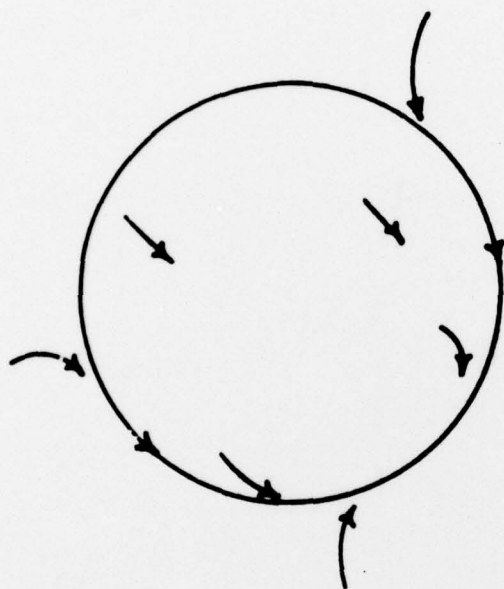
**FIGURE 4   FLOW LINES FOR INEQUALITY
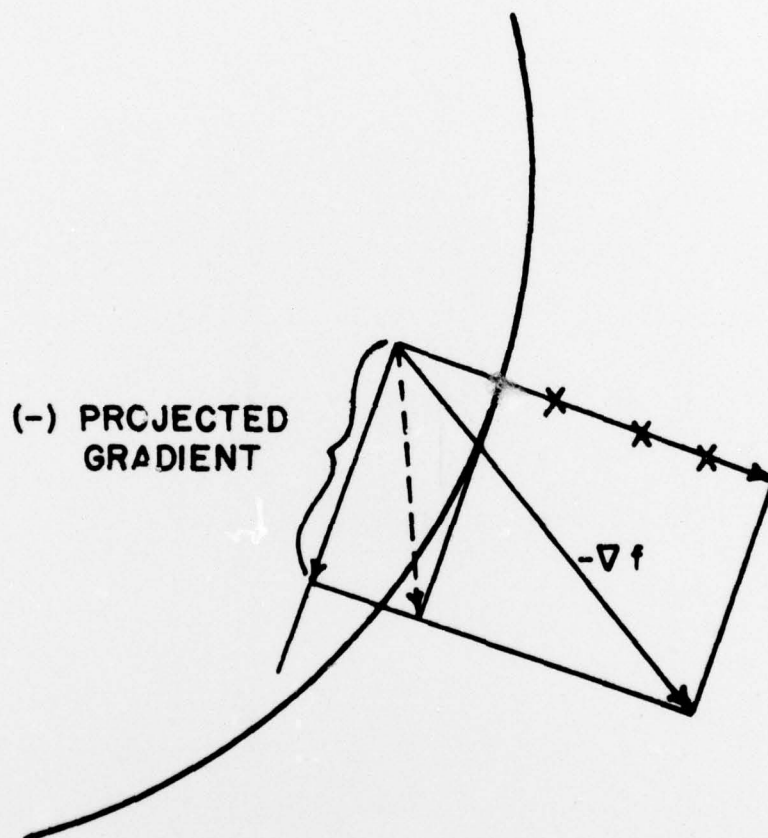CONSTRAINED PROBLEM**

(−) PROJECTED
GRADIENT

−∇f

FIGURE 5   FORCE RESULTING FROM
BOUNDARY REPULSION

$$E(x,y) = \begin{cases} -x+y & \text{IF} \quad -x^2-y^2+1 \geq \tau = .9 \\ -x/(x^2+y^2)^{.5} + y/(x^2+y^2)^{.5} \\ +[(x^2+y^2)^{.5}-1]^2 \quad \text{OTHERWISE} \end{cases}$$
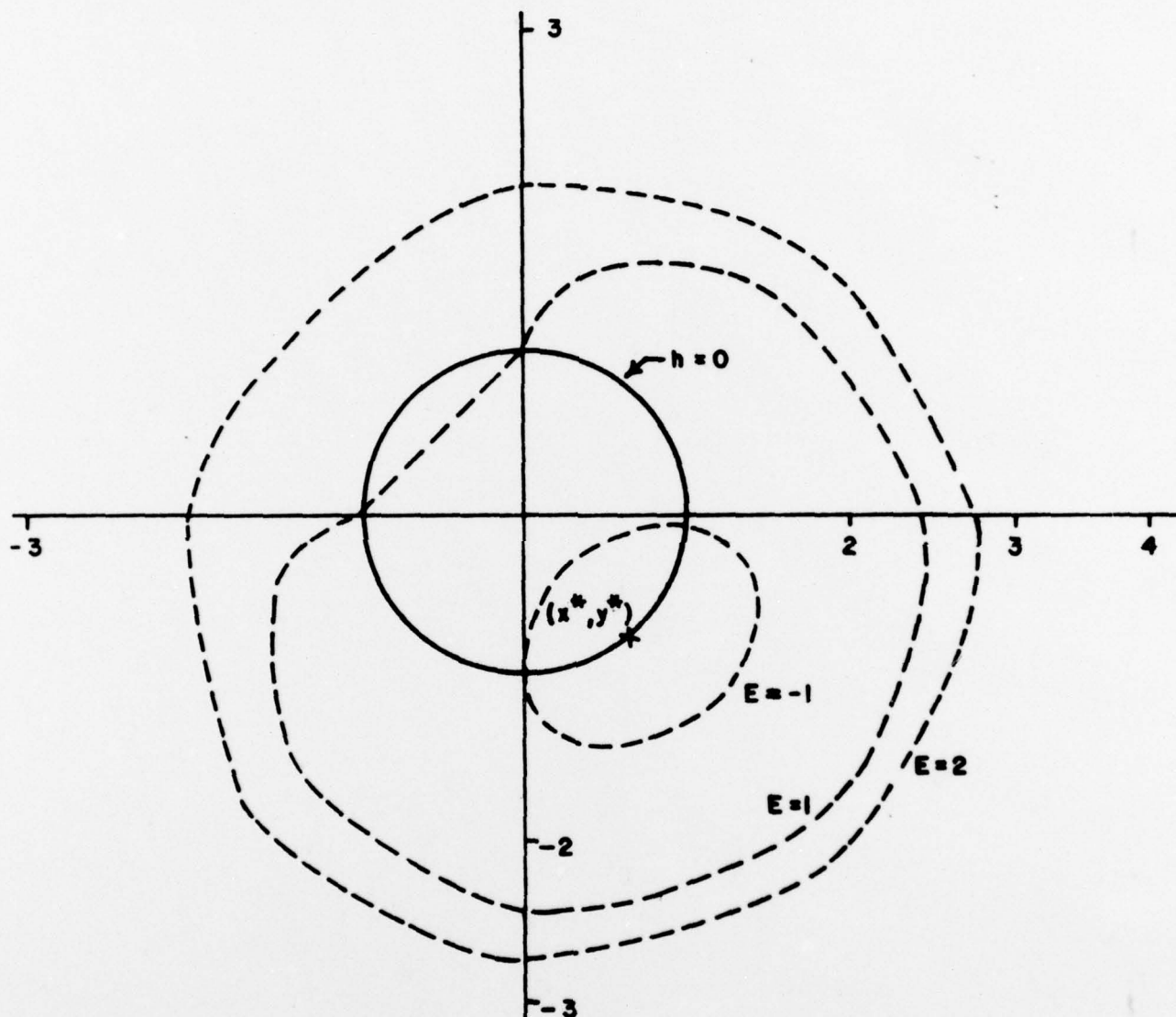

FIGURE 6   ISOVALUE CONTOURS FOR
INEQUALITY PROBLEM

scaling of the objective function and modifications must be made for the circumstances which can occur here but not in the equality case. There is also a serious combinatorial difficulty which arises. Which of the many subspaces defines the exact penalty function? In short, the inequality constrained problem poses many more difficulties than the equality constrained one.

Let $M = \{1,\dots,m\}$. There are 2^m subsets of M. Denote the ℓth subset by I_ℓ. By convention, let $I_1 = \{\phi\}$. Let $\tau > 0$ be a pre-assigned number which defines the "close" boundary points.

Also, define
$$S_\ell = \{y \mid g_i(y) = 0 , i \in I_\ell , y \in R\} . \tag{28}$$

The minimum value associated with S_ℓ is $E_\ell(x) = +\infty$ if $g_i(x) \geq \tau$ for some $i \in I_\ell$, or if S_ℓ is empty. Otherwise let

$$\mathscr{A}_\ell(x) = \min_{y \in S_\ell} \| x - y \| . \tag{29}$$

For any point x define $D_\ell(x) = \{d \mid x-d$ solves (29) above$\}$. Usually there is only one vector in $D_\ell(x)$. Let $d_\ell(x)$ be any vector from $D_\ell(x)$ such that

$$f[x-d_\ell(x)] = \inf_{d \in D_\ell(x)} f[x-d] . \tag{30}$$

Define
$$E_\ell(x) = f[x-d_\ell(x)] + q d_\ell^T(x)\, d_\ell(x) .$$

The exact penalty function associated with Problem (26) is

$$E(x) = \min_{\ell=1,\dots,2^m} \{E_\ell(x)\} . \tag{31}$$

This exact penalty function agrees with (5) when there are no inequality constraints.

Some isolvalue contours associated with Problem (27) are plotted in Figure 6. Here $\tau = .9$.

Theorems follow which relate the local minimizers of the exact penalty function to those of Problem (26).

__Theorem 8.__ Suppose f, $\{g_i\}$ are continuous. If $\bar{x}$ is a local unconstrained minimizer for the exact penalty function (31), then it is a local minimizer for Problem (26).

__Proof:__ Suppose $\bar{x}$ is a local minimizer for (31). Let ℓ be any index defining $E(\bar{x})$ (usually there is only one such index). Let $d_\ell(\bar{x})$ be some vector from $D_\ell(\bar{x})$ for which $E_\ell(\bar{x})$ is defined in (30).

Let ε be such that $0 < \varepsilon < 1$. Since $\bar{x} - \varepsilon d_\ell(\bar{x}) - (1-\varepsilon)d_\ell(\bar{x}) \varepsilon R$, it follows from the definition of $d_\ell[\bar{x}-\varepsilon d_\ell(\bar{x})]$ that

$$|| d_\ell[x-\varepsilon d_\ell(\bar{x})] || \leq (1-\varepsilon) || d_\ell(\bar{x}) || . \tag{32}$$

Also by definition

$$|| d_\ell(x) || \leq || \bar{x} - \left\{ [\bar{x}-\varepsilon d_\ell(\bar{x})] - d_\ell[\bar{x}-\varepsilon d_\ell(\bar{x})] \right\} || \tag{33}$$

$$\leq \varepsilon || d_\ell(\bar{x}) || + || d_\ell[\bar{x}-\varepsilon d_\ell(\bar{x})] ||. \text{ (from the triangular inequality)}$$

The last inequality is strict unless $d_\ell(\bar{x})$ is proportional to $d_\ell[\bar{x}-\varepsilon d_\ell(\bar{x})]$. Together (32) and (33) imply that

$$|| d_\ell[\bar{x}-\varepsilon d_\ell(\bar{x})] || = (1-\varepsilon) || d_\ell(\bar{x}) || .$$

Thus, $(1-\varepsilon)d_\ell(\bar{x}) \varepsilon D_\ell[\bar{x}-\varepsilon d_\ell(\bar{x})]$. $\tag{34}$

Because of (30).

$$f[\bar{x}-\varepsilon d_\ell(\bar{x})-d_\ell\{\bar{x}-\varepsilon d_\ell(\bar{x})\}] \leq f[\bar{x}-\varepsilon d_\ell(\bar{x})-(1-\varepsilon)d_\ell(\bar{x})] = f[\bar{x}-d_\ell(\bar{x})] . \tag{35}$$

Thus,

$$E[\bar{x}-\epsilon d_\ell(\bar{x})] \leq E_\ell[\bar{x}-\epsilon d_\ell(\bar{x})] \tag{(31)}$$

$$\leq f[\bar{x}-d_\ell(\bar{x})] + (1-\epsilon)^2 d_\ell^T(\bar{x})d_\ell(\bar{x})q \tag{(34) \& (35)}$$

$$\leq f[\bar{x}-d_\ell(\bar{x})] + d_\ell^T(\bar{x})d_\ell(\bar{x})q \tag{36}$$

$$= E(\bar{x}) .$$

If $d_\ell^T(x) \neq 0$, strict inequality holds in (36). Since $\bar{x}$ is an assumed local unconstrained minimizer for E, it follows that $E[\bar{x}-\epsilon d_\ell(\bar{x})] \geq E(\bar{x})$ for ϵ near one and therefore that $d_\ell(\bar{x}) = 0$. Thus, $\bar{x} \in R$, and also,

$$E(\bar{x}) = E_\ell(\bar{x}) = f(\bar{x}) . \tag{37}$$

Now, for any $x \in R$, $E_1(x) = f(x)$. Because $\bar{x}$ is an unconstrained local minimizer for E, there is a neighborhood $N(\bar{x},\delta)$ such that $E(x) \geq E(\bar{x})$ for $x \in N(\bar{x},\delta)$. Let $x \in N(\bar{x},\delta) \cap R$. Hence,

$$f(x) = E_1(x) \qquad (x \in R)$$

$$\geq E(x) \qquad ((31))$$

$$\geq E(\bar{x}) \qquad (x \in N(\bar{x},\delta))$$

$$= f(\bar{x}) . \qquad ((37)) . \qquad Q.E.D.$$

The importance of this theorem is that use of the exact penalty function does not introduce spurious candidates for local minimizers of the inequality constrained problem. The converse of this theorem is not true without modification; but this is not a serious computational problem. There are two partial converse theorems.

Theorem 9. Suppose f, $\{g_i\}$ are continuous functions. Suppose $\bar{x}$ is some local minimizer for Problem (26). Then there exists a value $\bar{q} > 0$ such that for $q \geq \bar{q}$, $\bar{x}$ is a local unconstrained minimizer for (31).

Proof: Assume the contrary, that there is a sequence $\{x_k\}$ where $x_k \to \tilde{x}$ such that

$$E(x_k) < E(\bar{x}) \leq f(\tilde{x})$$

for all k. Let $\ell(k)$ be an index defining $E(x_k)$.

Case (i). Without loss of generality assume that $\bar{x} \in S_{\ell(k)}$ for all k. It is clear that for any value of $q > 0$, the proof of Theorem 2 applies and a contradiction results.

Case (ii). Assume without loss of generality that $\bar{x} \not\in S_{\ell(k)}$ for all k. This means that there is an index i such that $g_i[x_k - d_{\ell(k)}(x_k)] = 0$ and $g_i(\bar{x}) > 0$. Thus,

$$\liminf_{k \to \infty} \|d_{\ell(k)}(x_k)\|^2 = \bar{v}_2 > 0 .$$

Let $\bar{v}_1 = \liminf_{k \to \infty} f[x_k - d_{\ell(k)}(x_k)]$.

Then, take $\bar{q}$ to be any value such that

$$\bar{q} > [f(\bar{x}) - \bar{v}_1]/v_2 .$$

It then follows directly that for k large

$$E(x_k) > f(\bar{x}) ,$$

a contradiction.   Q.E.D.

Theorem 10. Suppose f, $\{g_i\}$, $\{h_j\}$ are continuous functions. Suppose $\bar{x}$ is some local minimizer for Problem (26). Then there exists a value $\bar{\tau} > 0$ such that for all $0 < \tau \leq \bar{\tau}$, used in (28) $\bar{x}$ is a local unconstrained minimizer for (31) for every $q > 0$.

Proof: If τ used in defining is small enough, then there is a neighborhood about $\bar{x}$ such that for every x that neighborhood, $\bar{x} \in S_{\ell(x)}$ where $\ell(x)$ is an index defining $E(x)$. Thus, the Case (i) proof of Theorem 9 applies.

Q.E.D.

-23-

5. Fletcher's Exact Penalty Function (Inequality Case)

In [2] Fletcher extended his penalty function to the case when the problem is written with inequality constraints (26). As indicated in the development of the idealized exact penalty function in Section 4, there is a basic combinatorial problem that is not associated with the equality constrained problem. The problem is to determine, at any point x for which the penalty function is to be defined, which of the many constraints define the exact penalty function. The method given in Section 4 is theoretically valid, but possibly computationally prohibitive. Fletcher resolved this problem in the following way.

Consider the quadratic optimization problem:

for any point x ,

$$\underset{\delta}{\text{minimize}} \quad \frac{1}{2} q \ \delta^T \delta + f'(x)\delta$$

$$\text{subject to} \quad g(x) + g'(x)\delta \geq 0 \ . \tag{38}$$

Let $u(x,q)$ be a set of Kuhn-Tucker-Karush multipliers associated with the solution of the problem and let $\delta(x,q)$ denote the solution point. Since the objective function is strictly convex, the solution point (if one exists) is unique although, in general, the multipliers are not. The exact penalty function at x is, then,

$$R(x) = f(x) - u(x,q)^T g(x) \ . \tag{39}$$

When the multiplier vector $u(x,q)$ is not unique, the definition is ambiguous. Furthermore there are questions of the continuity of (39) and its levels of differentiability. These are taken up by Fletcher and under certain regularity assumptions he proves theorems concerning these matters.

To show that (39) is an approximation to (31), consider the following. Define $\mathcal{U}(x)$ to be the set of indices such that $g_i(x) + g_i'(x) \delta(x,q) = 0$,

i.e., the constraints active or binding at the solution of (38). Then the usual regularity assumption is that the vectors

$$\{g_i'(x)\}, \quad \text{for } i \in \mathcal{A}(x)$$

are linearly independent.

When this regularity assumption is satisfied, the multiplier vector $u(x,q)$ is unique. The components associated with the constraints whose indices are not in $\mathcal{A}(x)$ are equal to zero. Let $\underset{\sim}{g}(x)$ denote the vector of constraints with indices in $\mathcal{A}(x)$. The multipliers for these constraints are given by the formula (41) below. Then the exact penalty function takes the form

$$f(x) - f'(x)\underset{\sim}{g}'(x)^+\underset{\sim}{g}(x) + \frac{1}{2} q \underset{\sim}{g}(x)^T\underset{\sim}{g}'(x)^+\underset{\sim}{g}'(x)^{+T}\underset{\sim}{g}(x) , \tag{40}$$

because

$$\underset{\sim}{u}(x,q) = -\underset{\sim}{g}'(x)^{+T}\underset{\sim}{g}'(x)^+\underset{\sim}{g}(x)q + g'(x)^{+T}f'(x)^T \tag{41}$$

and

$$\delta(x,q) = \left[I - \underset{\sim}{g}'(x)^{+T}\underset{\sim}{g}'(x)\right]f'(x)^T/q - \underset{\sim}{g}'(x)^{+T}\underset{\sim}{g}(x) .$$

The obvious connection between this one and his equality exact penalty function can be made. This derivation is different from the way in which (20) was constructed. The difference between this one and the idealized exact penalty function (31) can be analyzed in the same way that the equality penalty functions were.

It is interesting to show how Fletcher could have obtained a similar penalty function which is closer but still essentially different from the minimum distance point of view.

Consider the problem

$$\underset{\delta}{\text{minimize}} \quad f(x-\delta) + \frac{1}{2}\delta^T\delta q \tag{42}$$
$$\text{subject to} \quad g(x-\delta) \geq 0 .$$

For any x, let $\delta(x,q)$ denote a solution point (local), and $u(x,q)$ a set of associated multipliers.

The exact penalty function is defined at x by

$$C(x) = f[x-\delta(x,q)] + \frac{1}{2}\delta(x,q)^T\delta(x,q)q \ . \tag{43}$$

The questions of whether or not this has spurious local minimizers and whether or not minimizer of the original problem are minimizers of this one are taken up next.

<u>Theorem 11</u>. If f, $\{g_i\}$ are continuous, and if $\bar{x}$ is local minimizer for (26), then $\bar{x}$ is a local unconstrained minimizer for the exact penalty function (43) above for any $q > 0$.

<u>Proof</u>: Pick ϵ small enough so that $\bar{x}$ is a global minimizer for $x\epsilon \left\{ x \mid \|\bar{x} - x\| \leq 2\sqrt{2\epsilon/q} \right\} \cap R$. Let $x \in N_1 = \left\{ x \mid \|x - \bar{x}\| < \sqrt{2\epsilon/q} \right\}$. Consider the neighborhood about 0 such that $D = \left\{ \delta \mid \|\delta\| \leq \sqrt{2\epsilon/q} \right\}$. Then for $\delta_B \in BND(D)$, $\frac{1}{2}\|\delta_B\|^2 q = \frac{1}{2} \cdot \frac{2\epsilon}{q} \cdot q = \epsilon$. Furthermore, for δ_B, $\|x - \delta_B - \bar{x}\| \leq \|x - \bar{x}\| + \|\delta_B\| \leq 2\sqrt{2\epsilon/q}$. Thus

$$f[x-\delta_B] \geq f(\bar{x}) \ .$$

Together, then,

$$f[x-\delta_B] + \frac{1}{2}\|\delta_B\|^2 q \geq f(\bar{x}) + \epsilon \ .$$

But for $\bar{\delta} = (x-\bar{x})$,

$$f[x-(x-\bar{x})] + \frac{1}{2}\|\bar{\delta}\|^2 q = f(\bar{x}) + \frac{1}{2}\|\bar{\delta}\|^2 q$$

$$< f(\bar{x}) + \epsilon \ .$$

Therefore the infimum is taken on at an interior point.

The converse theorem will now be proved.

<u>Theorem 12</u>. If f, $\{g_i\}$ are continuous functions, then any local unconstrained minimizer for the exact penalty function (43) for any value of q is a local minimizer for the constrained Problem (26).

<u>Proof</u>: Let $\bar{x}$ be a local unconstrained minimizer for (43). We first show that

$$\delta(\bar{x}, q) = 0 . \tag{44}$$

Assume that $\delta(\bar{x}, q) \neq 0$. Consider points along the ray connecting x to $\bar{x} - \delta(\bar{x}, q)$. The minimum value of (42) is clearly less than

$$f[x - \delta(x,q)] + \frac{1}{2}q(1-\epsilon)^2 \; ||\; \delta(x,q)\;||^2 < f(x-\delta(x,q)] + \frac{1}{2}q \; ||\; \delta(x,q)\;||^2$$

for ϵ small. This contradicts the assumption. Thus (44) must hold.

Let x be any point feasible to (26), and "close to" $\bar{x}$. Then

$$f(x) \geq G(x) \text{ (since } 0 \text{ is feasible to (42) when } x \in R)$$

$$\geq G(\bar{x}) \text{ (assumption that } \bar{x} \text{ was a local unconstrained minimizer)}$$

$$= f(\bar{x}) \; (\delta(\bar{x}, q) = 0) . \qquad\qquad \text{Q.E.D.}$$

REFERENCES

[1] FLETCHER, R. (1970). A class of methods for nonlinear programming with termination and convergence properties. Integer and Nonlinear Programming (J. Abadie, ed.) 157-175 North-Holland, Amsterdam.

[2] FLETCHER, R. (1973). An exact penalty function for nonlinear programming with inequalities. Mathematical Programming 5 (2) 129-150.